



Neural Networks Search for the **NUCLEAR NEEDLE** in a **HAYSTACK**

Neural networks trained to search for images, video, and text from open sources of information on the Internet offer Livermore's nuclear proliferation analysts the latest tool to find early evidence of rogue states engaged in clandestine nuclear activity.

A nuclear nonproliferation analyst searches for evidence that can help detect and characterize the status of a state's nuclear fuel cycle activities, including those that could assist in the clandestine development of nuclear weapons. In addition to information derived from classified intelligence or other confidential means, an increasingly relevant information source for

analysts is the rapidly growing body of information openly available on the Internet. A hypothetical example might be an image of a gas centrifuge, which can be used to enrich uranium, in an obscure technical newsletter published by a research institution.

Although the publication may be freely available on the Internet, the image of a centrifuge might be unlabeled, or the caption only identifies the names of the VIPs touring the facility. The article may be only one of millions of Internet-searchable links, in many publications and languages.

In that sea of information, how could the analyst find the article with this vital clue? How to sort through billions of bytes of data to get straight to the evidence?

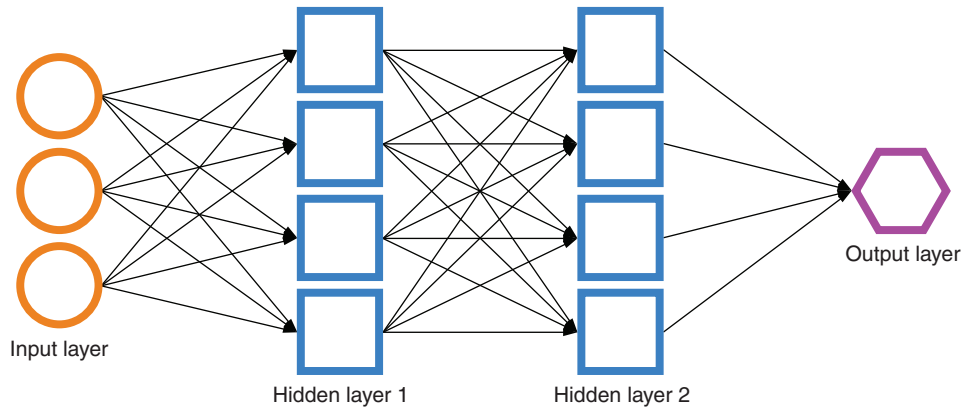
One of Livermore's mission research challenges is to develop innovative technologies to prevent, detect, counter, and respond to the use or threatened use of weapons of mass destruction. The Laboratory supports the work of the National Nuclear Security Administration (NNSA), other U.S. government agencies, and international bodies such as the International Atomic Energy

Agency (IAEA) to advance the nuclear nonproliferation mission. All the players in the nuclear nonproliferation community have a common goal: identifying early warning signs of proliferation—the development and spread of nuclear technologies for weapons—and a detailed, ongoing understanding of the status of extant nuclear weapons programs.

Livermore researchers are developing technologies that use data analytic capabilities to extract valuable nuggets of information from massive data streams to detect, characterize, and track nuclear weapons proliferation. Deep neural networks (DNNs) and machine learning offer a way to sift through these data streams for clues, which might be found in peer-reviewed scientific journals, local newspaper articles, patent applications, purchase orders for materials or equipment, and even job postings for nuclear-relevant skills.

LBANN Provides the Foundation

In 2015, a Livermore team led by machine learning researcher Barry Chen and high-performance computing (HPC)



Representation of a neural network showing input and output layers and hidden layers in between. Each shape is a gross approximation of a biological “neuron” that “fires” or activates based on inputs it receives. During the training process, each neuron learns input patterns on which to activate.

researcher Brian Van Essen developed the Livermore Big Artificial Neural Network (LBANN) Training Toolkit for accelerated training of large neural networks on HPC (See *S&TR*, June 2017, pp. 16).

An artificial neural network (ANN) is a machine learning model that “learns” a task by exposure to examples. ANNs are simplified mathematical models loosely inspired by biological brain structure. ANNs consist of individual “neurons” organized in multiple layers, each of which learns how to transform inputs into outputs optimized

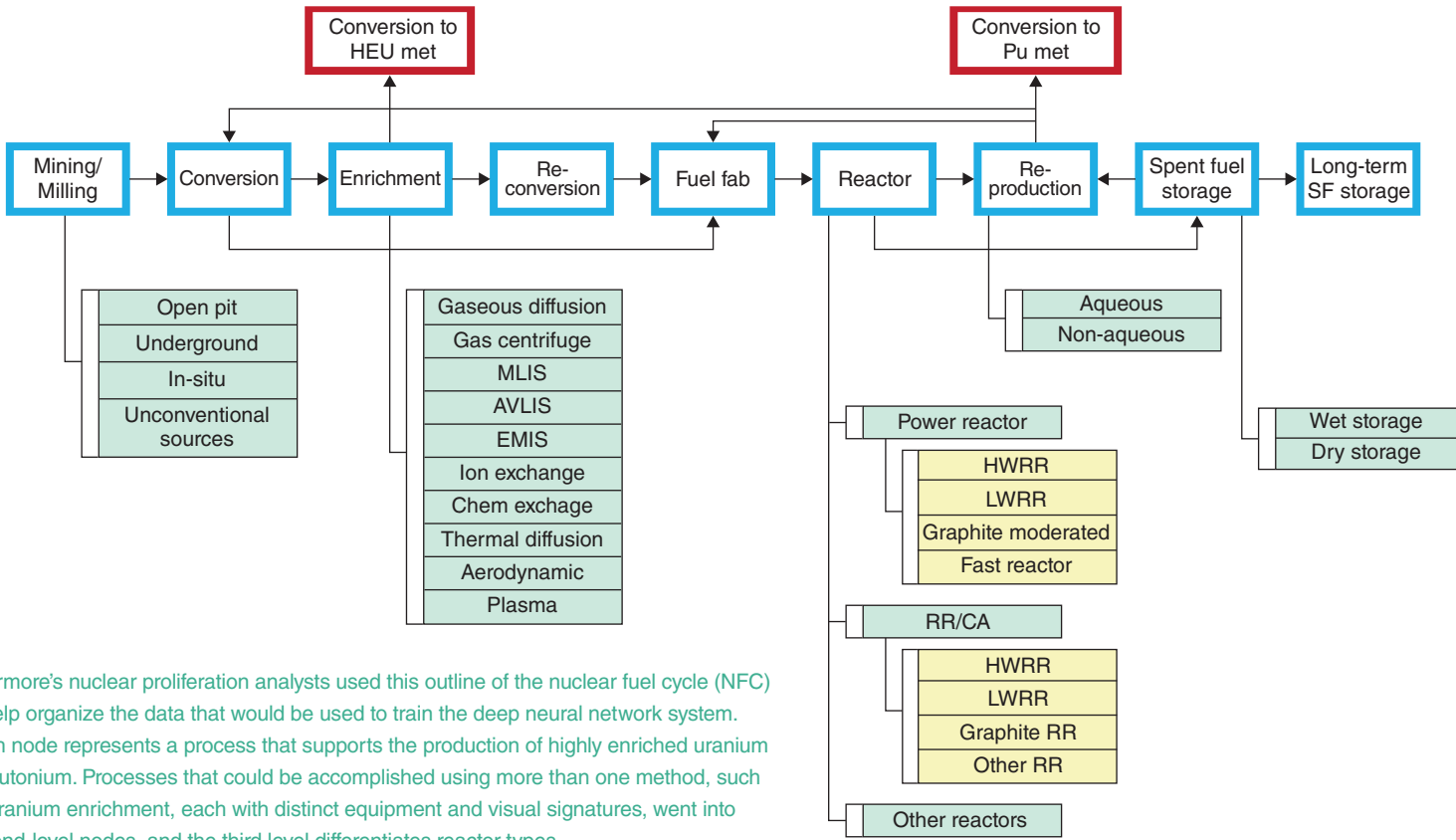
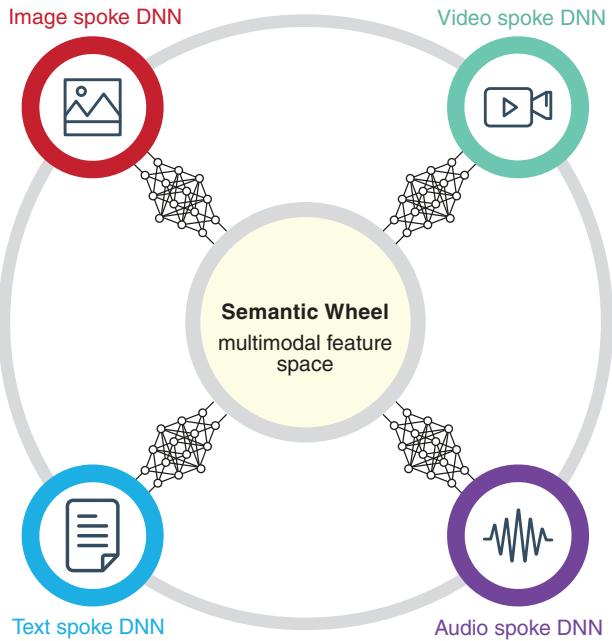
to achieve a task such as categorizing images into those with cars or trucks. Provided with multiple examples of images of cars and trucks, ANN layers progress through the input layer, one or more hidden layers, and finally the output layer, learning to successively detect meaningful

aspects from the original images to discriminate between cars and trucks.

Earlier layers “learn and remember” how pixels form edges, corners, and texture features. Intermediate hidden layers learn how these features form the bumpers, wheels, and doors of cars and trucks. Finally, the output layer combines everything to learn high-level details that distinguish car images from truck images, like the presence or absence of a trunk, truck bed, or hatchback. ANNs that have many layers between the input and output layers known as DNNs make it possible to model complex nonlinear mathematical relationships and solve large computational problems like image classification or object recognition.

After considering what kinds of problems a neural network could solve among the Laboratory’s missions, Chen and his collaborators recognized a need in nuclear proliferation analysis. “The analysts face a deluge of data. Retrieval of useful information is the challenge,” says Chen. “There’s too much information for any one person to sort through.” Chen

The Semantic Wheel is the conceptual idea behind the multimodal deep neural networks (DNN). Data contained in each mode’s spoke for text, audio, image, and video data are mapped into a multimodal “feature space” or mathematical abstraction, in which all data types representing a concept such as a gas centrifuge are mathematically in proximity.



Livermore’s nuclear proliferation analysts used this outline of the nuclear fuel cycle (NFC) to help organize the data that would be used to train the deep neural network system. Each node represents a process that supports the production of highly enriched uranium or plutonium. Processes that could be accomplished using more than one method, such as uranium enrichment, each with distinct equipment and visual signatures, went into second-level nodes, and the third level differentiates reactor types.

imagined that a neural network-based data retrieval system could make life easier for analysts by doing most of the sorting for them if the network could be trained to recognize and retrieve useful information. “There is a lot of open-source data that we’re leaving on the table because we can’t manually process it,” says Yana Feldman, associate program leader for international safeguards in Livermore’s Global Security Principal Directorate.

Learning Other Languages

Discussions with the Laboratory’s nuclear proliferation analysts led to a research project: “Large-Scale Multimodal Deep Learning for Nuclear Nonproliferation Analysis” funded by the Laboratory Directed Research and Development (LDRD) Program, begun in 2017. The three-year project goal was to deliver a system that allows analysts to retrieve multimodal open-source data such as text, image, audio, and video relevant to nuclear proliferation.

Developing the system required that the computational engineers, data scientists, and nonproliferation analysts assigned to the project learn new ways of working together, and even new ways of talking to each other. When the neural network experts sat down with the nonproliferation experts, they had to develop a shared lexicon to advance the project’s goals. Even a seemingly simple word like “data” had different definitions for different groups. “Data is a vague term,” says Brenda Ng, Livermore data scientist and machine learning group lead, “For us, ‘data’ is *labeled* input as well as the target output. To a proliferation analyst, ‘data’ is any video, text, or image. We worked with the analysts to understand how to categorize and label the data. That way, we could derive labeled data that the model can use to learn to map the data from text into specific categories.”

“We had to get enough labeled data to test the system so that we could give

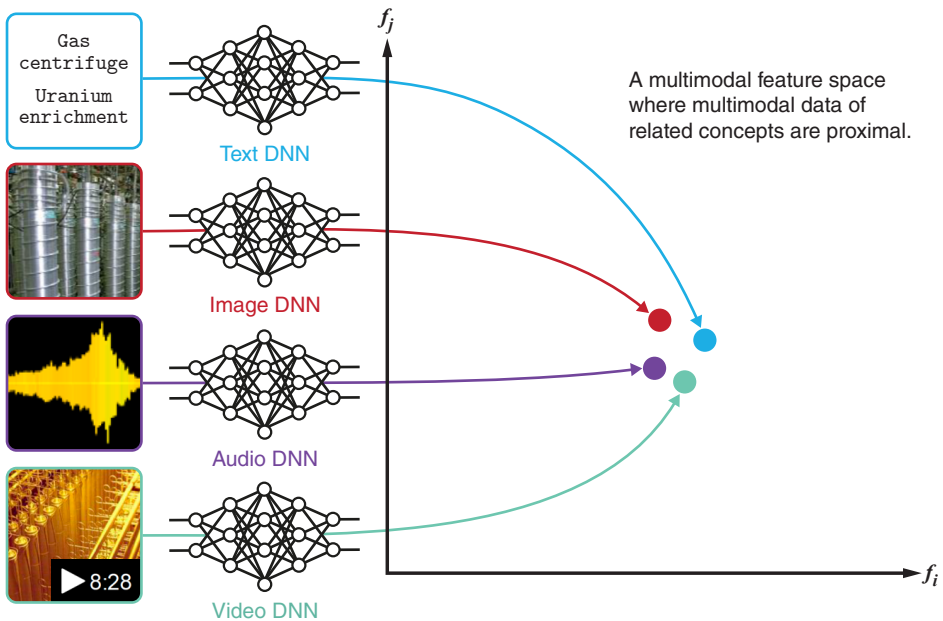
the analysts something that would work,” says Carmen Carrano, computational engineer at Livermore who works on image processing, machine learning, and video analytics and developed and tested the software codes for the system. “Coming up with the language for captions or tags for the images was a challenge. The analysts would give us sentences describing an image, but we needed to know which part of the sentence described what we’re actually looking at.” Ultimately, persistence paid off. “Lots of iteration was required,” says Feldman, “Relabeling, re-annotating videos, and so on. We found that working closely with the machine-learning specialists was an important factor in making this project a success.”

Neural Networks for Nonproliferation

“People might ask, aren’t companies like Google doing unlabeled image-video search retrieval? To some extent yes, they are. If you’re looking for images, you can perform a keyword search to find images

labeled with text descriptions. But for unlabeled images, keyword search doesn't work. That's why we decided to use neural networks to help us index unlabeled data, and more importantly, specialize the way the neural networks index data related to nuclear technologies. This lets us find images and video about specific nuclear technologies that do not have text descriptions," says Chen.

Central to the system, the team created the "Semantic Wheel." Each "spoke" of the wheel represents a data modality: text, image, audio, video. "Neural networks are a great way to index and organize data for easy searching," says Chen. "Data of different modalities would be projected into a classification learning problem with feature vectors called "a joint feature space," such that the distance between conceptually related data is small. This index provides a search mechanism for the analyst." For example, images and video of gas centrifuges, and text containing the words "gas centrifuge," would all reside "near" each other in a feature space calculated by the neural network. Chen describes the index as "an automated Dewey decimal system."



The team conducted the work in two phases. First, they trained a DNN to understand each data modality separately: one to find text, one to find relevant images, and so on. Each trained modality forms one spoke of the Semantic Wheel. Then, the team had to figure out how to assemble the Semantic Wheel's individual modality spokes into a multimodal feature space.

To begin, the analysts used a simplified nuclear fuel process model to train the DNNs. Each node represented a step in the process that supports the production of highly enriched uranium or plutonium. They also subdivided processes that use more than one method, such as uranium enrichment, each with distinct equipment and visual signatures, into second-level nodes, and then, where appropriate, added a third level to provide additional discrimination, among reactor types, for example.

Getting the Text Spoke Up & Running

To develop the text spoke, the team first gathered as much data as possible from freely available internet sources using automated tools. "We

had to develop methods to clean up the information, for example, to distinguish watermarks in PDF files from relevant information," says Computational Engineering Division machine learning scientist Sam Nguyen. The other big challenge was to adapt existing natural language processing models to the task of finding what proliferation analysts were looking for. The team needed to develop a compressed, numeric representation, or an "embedding," specific to nuclear nonproliferation phenomena. "To 'embed' means to take a data object and map it to a numeric representation so that this data object can be integrated into a neural network for further analysis," says Ng.

Industry search engine and social media giants have published embeddings for text, but the Livermore team needed to develop embeddings specific to the nuclear proliferation-related content they monitor. "At the time, ours was one of the earliest groups that trained our own models to learn customized embeddings for our application," says Ng, "Now, this is becoming more common. For example, the biomedical community has come up with BioBERT, which is their customized embedding for medical texts."

"Entity reconciliation" emerged as another success of the project. Ng's team tackled this challenge using context-sensitive models. "Analysts may refer to one thing in many different ways," she says, "So mapping different names to the same concept is a challenge. But in managing this, we were able to deliver a system to analysts that went beyond key word retrieval, which is really remarkable."

A system of DNNs that maps text, imagery, audio, and video into a multimodal feature space. Conceptually related data are "close" to each other. Here, text tags, an image, an audio file, and a video related to a gas centrifuge enrichment process are mapped to nearby points in the multimodal feature space.

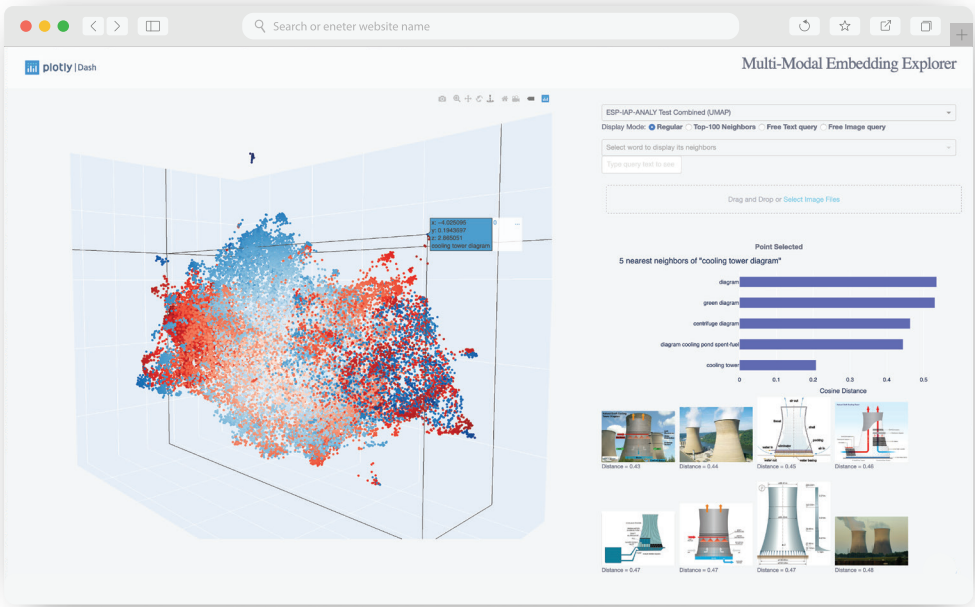
Training Neural Networks to Index

To train the DNN, the team gathered images and schematic diagrams of such technologies as uranium gas centrifuges, flow-forming machines, reactor fuel elements, cooling towers, spent fuel pools, reactor cores, hot cells, and uranium cylinders from a variety of open sources including U.S. Department of Energy and IAEA publications, Wikipedia, and various news sources.

The team used both labeled and unlabeled images and video to train the DNNs. The goal was to map raw unimodal data into a feature space where similar data are close to each other. The DNN would project these labeled images, for example of gas centrifuges, in mathematical proximity to one another in the feature space. The learning process generates n-dimensional vectors of numerical features that represent a target object called "unimodal feature vectors," which mathematically represent individual images, text excerpts, or video clips of the targeted object.

The second phase of the project was to train another DNN to map each of the unimodal feature vectors into a multimodal feature space that maps and aggregates conceptually related data to nearby locations. After the training, the feature vectors representing all text, images, and video of gas centrifuges, for example, are in close proximity within the multimodal DNN's feature space. Once this learning process is sufficiently advanced, the analyst can execute various searches using text to search for images, an image to search for like images, text to video, and image to video.

The video spoke presented unique data curation and computational modeling challenges. The information conveyed by a video is often greater than the individual frames. As a result, analyst annotations often expressed abstractions not tied to any particular frame or region of interest, such as the outcome of a process. Therefore, the video team and analysts



Data representing text, images, and video of like items, in this case cooling towers, are mapped into a feature space in close proximity to one another.

worked together to develop new ways to curate and label video data for training the video spoke. On the work's computational modeling aspect, engineer Doug Poland, who works on computer vision, video analytics, and machine learning, says, "Video presents a fundamental problem that has not been completely solved. We are developing a new framework based on spatial and temporal modeling capabilities of the human brain to better capture the complexities of video scenes."

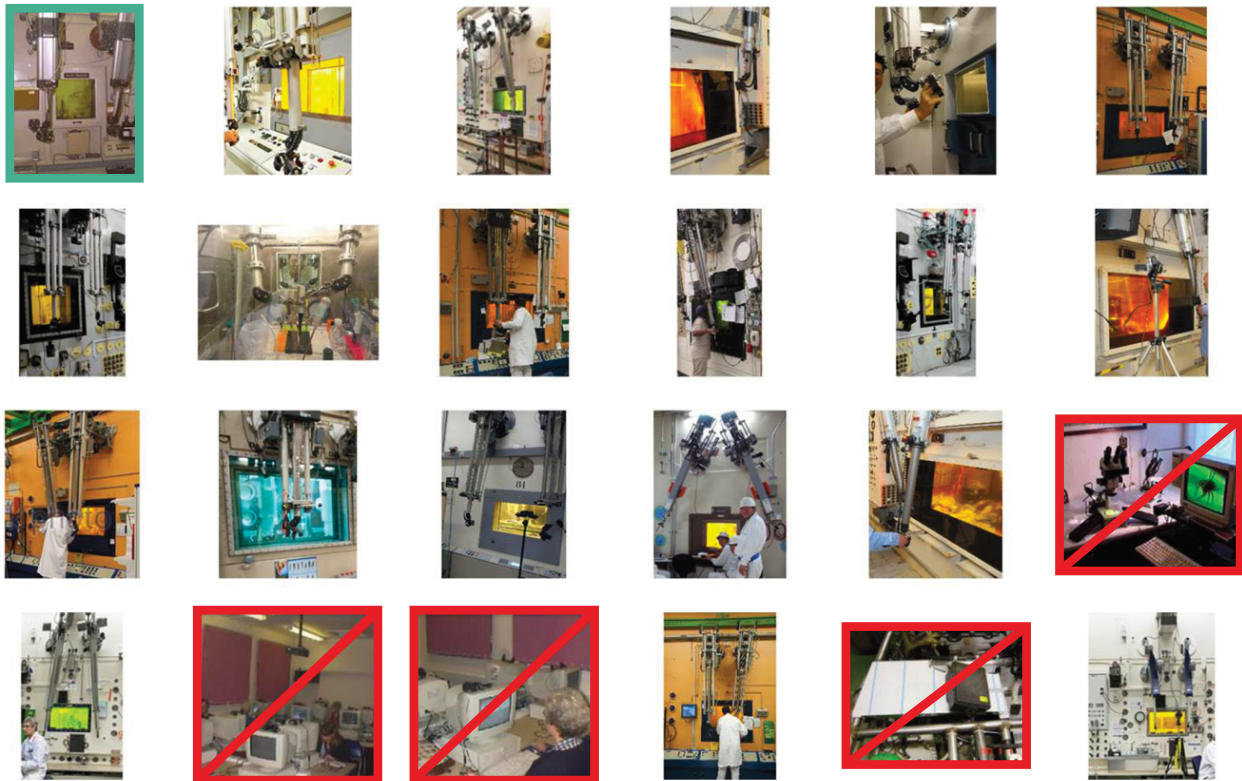
Unimodal Learning Saves Effort Later

One of the innovations emerging from this work is the insight that training the unimodal spokes first saves learning effort later. "You can recycle a lot of relationship data from the unimodal data so that you don't need as much multimodal data for training," says Chen. "Consider, for example, the concepts of airplanes and clouds. In images you'd likely see airplanes flying through clouds or flying above clouds. This sort of relationship also manifests itself in text, so when training

the multimodal DNN, our system will recycle these pre-learned unimodal relationships in the multimodal feature space." In this way, training the neural networks of the unimodal spokes first reduces some of the data required to train the multimodal DNN.

"The lack of truly large-scale aligned, multimodal datasets presents a challenge in multimodal training," says Jaeyoung Choi, a graduate student at the University of California, Berkeley-affiliated International Computer Science Institute. "The individual datasets in each modality were not always big enough to train the individual spokes—data collection has been a bottleneck in much of machine learning's development. One potential solution is to combine multiple datasets for training." Several issues needed to be addressed such as different semantic coverage between datasets and certain concepts being over or under-represented quantitatively—a lot of text and fewer images for a given element of the nuclear fuel cycle, for example.

Seed image



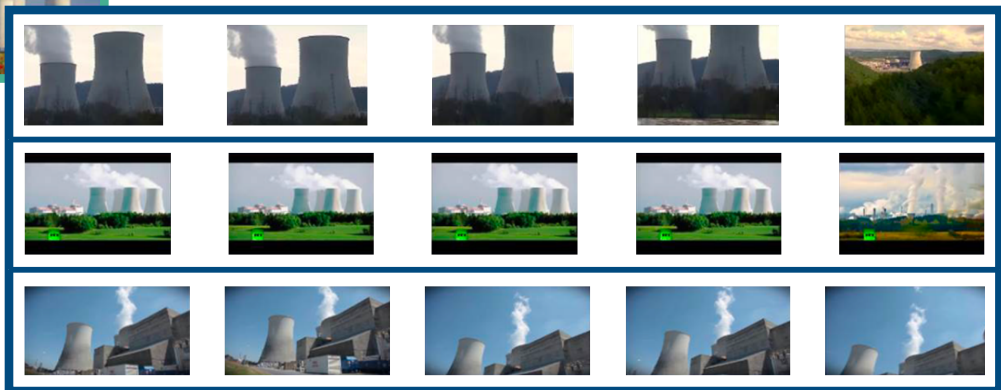
19 accurate hits out of 23

An example of an image-to-image search conducted by the multimodal DNN using an image of a hot cell, a type of chamber with a window and mechanical aids to grip objects, used to perform work with radioactive materials. Nineteen out of 23 found images were correctly identified as hot cells.

Seed image



Top 3 matching video and frames



An image-to-video search using a frame grab of a nuclear power plant's cooling tower, and the top three search results.

“The overall question was, ‘Can we combine multiple datasets and still get a good alignment so that similar multimodal data were close to each other in the hub?’ We had two objectives in developing training methods,” says Choi. “We did not want to worry about the different characteristics of the datasets. Second, we wanted to make the methodology scalable and efficient.” To address the first problem, multitask learning was integrated into the framework with “joint loss weight optimization.” This method treated multiple datasets as if they were one large dataset. Uncertainty-based weighting compensated for the loss of matched data as the multimodal dataset was scaled up by combining smaller datasets. During the model training process, these “loss weights” were jointly optimized with other model parameters. The Livermore work was the first to use uncertainty-based weighting to handle the loss-with-scale-up issue between multimodal data sets.

For the second problem, Choi developed a strategy of two-stage, shared representation optimization. In this process, each modality is first optimized individually—text or image or video. During the next step of intermodal optimization, the model’s unimodal semantic structure is transferred to a joint semantic space (text, images, and video together). In the joint space, paired data might be imperfectly matched. To deal with this problem, the model uses a “bidirectional quadruplet loss function,” so-called because it takes two pairs of aligned data as the input. The model now jointly optimizes the cross-modal semantic relationship of the pairs, compensating for imperfect alignment of data between modalities. This process was crucial to the model’s ability to learn a discriminative joint semantic structure—in other words, learning how to find the object the analyst is looking for whether it is text, image, or video. The result was a robust DNN

whose performance showed significant improvement compared to training methods reported by other researchers.

Working System Delivered

The system the machine-learning team demonstrated to the analysts can conduct several types of multimodal searches. The input of a sample image of a hot cell successfully returns unlabeled images of hot cells from among tens of thousands of possibilities. Text-to-image and text-to-text searches also show a high degree of accuracy. Image-to-video, text-to-video, and video-to-text searches show promising results. The video spoke, which is still in development, has demonstrated success at executing an image-to-video search, identifying cooling towers within video frames from a sample cooling tower image.

Data scientist Steven Samson developed the user interface and the search index that stores, delivers, and displays target information to the user interface as quickly as comparable commercial search engines. “We wanted to make sure that the interface was responsive—delivering with a sufficiently rapid turnaround,” says Samson. “This system gives me the capability to quickly find things that I would not have been able to find manually,” says Feldman. “It could potentially alert me to something that might interest me—a two-second frame in a two-hour video. Without it, I’d have to watch the entire video, and with hundreds of hours of new video being added to the internet every minute, that’s just not practical. It gives us the ability to process more data than we ever could before.”

The team plans to further develop partnerships to improve the system by advancing the video spoke and adding an audio spoke, as well as other types of data such as patent schematics. “We would also love to customize the system to meet the needs of individual analysts. Having more analysts in the loop to provide feedback on retrieval results will help the system learn from its mistakes and improve performance,” says Chen.

A new multilaboratory effort funded by the NNSA’s Defense Nuclear Nonproliferation R&D called Advanced Data Analytics for Proliferation Detection (ADAPD) is also looking to leverage the system to help with its own mission. ADAPD brings together Livermore, Los Alamos, Oak Ridge, Pacific Northwest, and Sandia national laboratories to develop a global-scale, real-time capability to detect, locate, and characterize low-profile proliferation. Eddy Banks, ADAPD’s principal investigator at Livermore, says, “We need a predictive capability to detect the steps that a nation might take as it moves toward weapons development.”

Detecting seismic signals from a nuclear test is too late—by then, the proliferator already has a weapon. ADAPD is looking to detect signals from earlier activities like hiring people with particular nuclear weapons engineering expertise, orders of relevant equipment, or an increase in traffic around a targeted facility. Detecting these signs provides the nonproliferation community with the evidence it needs to mobilize international efforts to intercede. “This is where the work of Barry Chen’s group is important to ADAPD,” says Banks. This partnership between human minds and deep neural networks to find and interpret evidence promises to help the international community reduce the dangers of the spread of nuclear weapons.

—Allan Chen

Key Words: Advanced Data Analytics for Proliferation Detection (ADAPD), artificial neural network (ANN), information retrieval, bidirectional quadruplet loss, deep neural network (DNN), semantic embeddings, joint loss weight optimization, Laboratory Directed Research and Development (LDRD) Program, Livermore Big Artificial Neural Network (LBANN), machine learning, multimodal deep learning, natural language processing, nuclear fuel cycle (NFC), nuclear proliferation, nonproliferation, open-source information, Semantic Wheel.

For further information contact Barry Chen (925) 423-9249 (chen52@llnl.gov).